Day 2, Morning
Statistical Intervals
Hypothesis Tests

Statistical Inference

- Population and samples
- Remember, to use sample data for inference, needs to be representative of population *for the question(s) of interest.*
- Some definitions:
  - A **parameter** is a number associated with a population. Assumed to be fixed but unknown.
  - A **statistic** is a number computed from a sample. Changes with each new sample.

Statistical Inference: Intervals

- Estimate a parameter using sample data.
  - Point estimate: A single number.
  - Interval estimate (aka **confidence interval**): An interval of values we are fairly confident covers the true population parameter.
- Find an interval that is likely to cover a specified percentage of *new* values from the same population
  - This is a **prediction interval**.

Statistical Inference: Hypothesis Tests

- Test whether a parameter = a specific value, versus either *not equal*, *greater than*, or *less than* that value.
- Special case: In regression, test whether the slope of the line = 0, meaning there is no linear relationship between x and y.
- Compare populations, for instance to see if means for several populations are equal.

Parametric and Nonparametric

- Parametric methods (bit of a misnomer):
  - Based on assuming a particular underlying population distribution, usually "normal."
  - Can sometimes be used even without that assumption, for large samples.
- Nonparametric methods:
  - Can be used without assuming a distribution.
  - Often not as "powerful" as parametric methods.
- When in doubt, safer to use nonparametric method

Confidence Intervals

- A **parameter** is a population characteristic – value is usually unknown. We estimate the parameter using sample information.
- A **statistic,** or **estimate,** is a characteristic of a sample. A statistic estimates a parameter.
- A **confidence interval** is an interval of values computed from sample data that is likely to include the true population value.
- The **confidence level** (often .95) for an interval describes our confidence in the procedure we used. *We are confident* that most of the confidence intervals we compute using our procedure will contain the true population value.

Hands-On Activity

- Applet to demonstrate confidence interval concepts

http://www.rossmanchance.com/applets/New Confsim/Confsim.html

- Details on activities handout
- Note that on average, about 19 out of 20 of all 95% confidence intervals should cover the true population value.

Example: pH of Rain in Davis
1980 to 2009; what is population mean?

Sample size = 30, reasonably bell-shaped
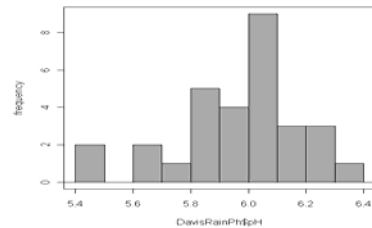
Confidence interval for mean

- Assume population is close to normal.
- Also works if n is large (30 if no outliers or major skew, perhaps 50 or so if more skewed)
- Based on knowledge of *sampling distribution* of the sample mean.
- Sampling distribution gives range and probability distribution for all possible sample means if infinite number of samples of size n taken from the population.

Notation for Means, Quantitative Variables

- Take sample of size *n* from a population and measure a quantitative variable.
- Notation for Population (uses Greek letters):
  - μ = mean for the *population* of measurements.
  - σ = standard deviation for the *population*.
- Notation for Sample:
  - $\overline{x}$ = *sample* mean for a sample of *n* individuals.
  - s = *sample* standard deviation for the sample.


Sampling Distribution of the Sample Mean

- Suppose the distribution of possible values is normal with mean μ and standard deviation σ and/or n is large.
- The distribution of possible values for the sample mean $\overline{x}$ is normal with mean μ and standard deviation $\dfrac{\sigma}{\sqrt{n}}$

Example: pH values in Rainfall

- Suppose *individual* pH values in annual rainfall in a certain area are from a normal distribution with $\mu = 5.6$ and $\sigma = 0.5$.
  - From Empirical rule, almost all *individual* values are with $3(0.5) = 1.5$ of mean,
  - In range 5.6 +/- 1.5, or 4.1 to 7.1
- *Mean* pH for n = 25
  - normally distributed with mean $\mu = 5.6$ and standard deviation = $\sigma/\sqrt{n} = 0.5/5 = 0.1$
  - *Mean* almost always in the range 5.6 +/- 0.3, or 5.3 to 5.9.

Slide 1

Comparing individual pH values with Mean pH for n=25

StDev
— 0.5
— 0.1

Mean of 25 pH values

Individual pH values

pH
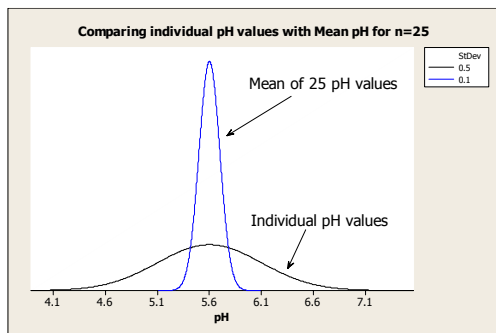
Also, from Empirical Rule, about 95% of *sample means* in range 5.4 to 5.8.

Confidence Interval for Mean μ

- Suppose we *know* population s.d. σ = .5, and we have *one* sample mean $\overline{x}$ = 5.67.
- For *n* = 25, 95% of the time $\overline{x}$ is within 2(.1) = .2 of the true mean μ.
- If this is one of those times, then we know μ and $\overline{x}$ are no more than 0.2 apart.
- So, we guess that μ is in the interval 5.67 ± .2 or 5.47 to 5.87.
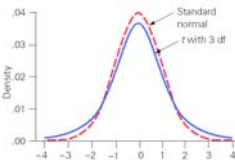- We have just found a *95% confidence interval* for μ.

Student's *t*-Distribution: Replacing σ with *s*

Dilemma: we generally don't know σ. Using *s* we have:

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}} = \frac{\sqrt{n}(\overline{x} - \mu)}{s}$$

If the sample size *n* is small, this standardized statistic will not have a *normal* distribution but rather a *t-distribution* with *n* – 1 *degrees of freedom* **(df).**

Standard normal
t with 3 df

General Formula, C.I. for a Mean

$$\overline{x} \pm t^* \times \frac{s}{\sqrt{n}}$$

Explanation of the pieces:
$\overline{x}$ = sample mean
t* = value from "t distribution"
s = sample standard deviation

Finding the *t*-multiplier

- R Commander:

  Distributions → Continuous distributions → t distribution → t quantiles
  – Probabilities: For 95% C.I., use .025
  – Degrees of freedom = *n* – 1
  – Lower tail
  – Gives negative of the t-multiplier
  – Ex: .025, 25, lower tail → −2.059539, multiplier ≈ 2.06

Using R Commander to find a confidence interval for a mean directly

- R Commander links tests and confidence intervals.

  Statistics → Means → Single sample t-test
- Give desired *confidence level* (most common is .95, but sometimes use .90 or .99)
- Ignore remaining options for now (they are for hypothesis tests).

Activity: Find 90% C.I. for mean pH for Davis rainfall

- Import data from Excel: *DavisRain*
- Find a stem and leaf plot. Does it look approximately normal?
- Find a 90% confidence interval for the mean. (Use pH, change level to .90)

Statistics → Means → Single sample t-test

- Interpret the interval. What mean is it for?

Day 2, Morning, Slide **19**

Results should look like this

```
1 | 2: represents 0.12
 leaf unit: 0.01
            n: 30
   54 | 79
   55 |
   56 | 1
   57 | 09
   58 | 26788
   59 | 1578
   60 | 34456678
   61 | 0133
   62 | 247
   63 | 3
```

```
90 percent confidence
  interval:
  5.906136 6.036531
sample estimates:
mean of x
 5.971333
```

Day 2, Morning, Slide **20**

Suppose lognormal is better model

- Note that for normal distribution, mean = median. So, a C.I. for the mean is also a C.I. for the median.
- Not true for lognormal distribution (mean ≠ median). We will find C.I. for median, not mean.
- Basic idea: Transform data using $y = \log(x)$, find C.I., transform back.

Day 2, Morning, Slide **21**

Rationale

- Suppose variable x has lognormal distribution.
- Then $y = \log(x)$ has normal distribution.
- *mx* = median(x), value with 50% of x below it.
- *my* = median(y) = value with 50% of y's below it.
- *my* = log(*mx*), i.e. can find 50% point and then take log *or* take logs then find 50% point.
- This is *not* true for the means. Mean of log(x) is *not* the same as log of mean(x).

Day 2, Morning, Slide **22**

Rationale, continued
Want C.I. for median(*x*) = *mx*

- Find 95% C.I. for *mean* of y, same as *median* of y, *my*. (Remember, y's are normal.)
- Remember that *my* = log(*mx*)
- Suppose interval is L to U. We are 95% confident that L < *my* < U, or L < log(*mx*) < U
- So 95% confident that exp(L) < *mx* < exp(U)
- So 95% confidence interval for *mx* exp(L) to exp(U)

Day 2, Morning, Slide **23**

To Summarize

- Transform *x*'s by using $y = \log(x)$ (Natural log)
- Find 95% confidence interval for mean (median) of *y*, using *t* interval as before
- Transform endpoints of interval using exponential function.
- Result is a 95% confidence interval for the median of *x*.

Day 2, Morning, Slide **24**

Activity: Benzene data

- Import data from Excel: *Benzene*
- Look at histogram to see right skewed.
- Create new variable, logBenz

Data → Manage variables in → Compute new...
  logBenz is created as log(Benzene)

- Find 95% confidence interval for mean of logBenz. Should get 4.29842 to 5.85981
- Use calculator to find exp(4.29842) to exp(5.85981) is 73.583 to 350.658

Benzene example, continued

- 95% confidence interval for population *median* benzene is 73.583 to 350.658.
- If we had used normal model assumption on original benzene data, 95% confidence interval for the *mean* would be 133.9483 to 520.0267.
- Remember that mean is 327, median is 190, so they are very different.

Activity

- Import the Saddle River data from Excel file.
- Investigate the shape of the "Flow" variable.
- Compare the sample mean and sample median.
- Get a 95% confidence interval for the mean.
- Transform and get a 95% confidence interval for the median that way.
- Results: C.I. for mean: (1447.787, 1906.644)
  C.I. for median: exp(7.134264, 7.413504), which is (1254.21, 1658.23)

Prediction Interval

- A *prediction interval* is an interval predicted to include a new observation from the same population with specified probability.
- Example: A 90% confidence interval for the *mean* pH for Davis rainfall is 5.91 to 6.04.
- What interval do we predict 90% of *future individual values* to fall in, assuming the past and future represent the same population?

Nonparametric Prediction Interval

- Suppose there are 30 observations. What is the probability that a new, 31st observation, will fall between the minimum and maximum of the first 30 observations?
- Probability that the 31st observation will be the largest is 1/31. Similarly, probability that it will be the smallest is 1/31. So probability that it will not be largest or smallest is 29/31.
- Thus, the interval (min to max) is a 29/31 or 93.5% *prediction interval*.

Nonparametric P.I., continued

- In general, if there are n observations, then (min to max) is a $(n-1)/(n+1)$ prediction interval.
- Reasoning can be extended. For a $(1-\alpha)$ P.I., find c $=(\alpha/2)(n+1)$. Then the interval is the the c-th smallest to c-th largest observation.
- Ex: 90%, $\alpha = .10$. For $n = 30$, c = (.05)(31) = 1.55. Interpolate, or roughly, use 1.5^{th} smallest and largest.
- Davis pH data: Lower end, halfway between 1st and 2nd smallest (5.47+5.49)/2=5.48. Upper end, halfway between 6.27 and 6.33, or 6.30.

Hands-On Activity

- Find a 90% prediction interval for a new value based on the San Francisco effluent nickel data.
- There are 39 observations
- Lowest 3 observations are 1.9, 2.0, 2.1
- Highest 3 observations are 3.7, 3.9, 4.4

Parametric P.I., assuming normal

- If we knew μ and σ, then 95% P.I would be about $\mu \pm 2\sigma$.
- More complicated otherwise.
  - Because of estimating μ, need to add same variability we use to get C.I. for μ .
  - Because of estimating σ, need to use t* as multiplier instead of 2.

Parametric P.I., continued

Prediction interval is

$$\bar{x} \pm t^* s \sqrt{1 + \frac{1}{n}}$$

where t* is chosen so the area above it is $\alpha/2$ in a *t*-distribution with degrees of freedom = $n - 1$. There is no easy way to do this in R Commander, but see next slide.

Using C.I. to find P.I.

Recall, a confidence interval is $\bar{x} \pm t^* \frac{s}{\sqrt{n}}$
Prediction interval is

$$\bar{x} \pm t^* s \sqrt{1 + \frac{1}{n}} \ \Rightarrow \ \bar{x} \pm t^* s \sqrt{\frac{n+1}{n}} \ \Rightarrow \ \boxed{\bar{x} \pm t^* \frac{s}{\sqrt{n}} \sqrt{n+1}}$$

So if you have a C.I. for μ:
- Find the "half-width" (part after the ± sign)
- Multiply it by $\sqrt{n+1}$ to get half-width for P.I.
- Add and subtract that half-width to $\bar{x}$

Davis pH Example

- R Commander gave 90 percent confidence interval as 5.906136 to 6.036531and mean of x as 5.971333.
- From these, easy to see half-width is .0652.
- $(.0652)\sqrt{31} = .3630$
- So a 90% P.I. is $5.971 \pm .363$ or 5.608, 6.334
- Nonparametric version was 5.55 to 6.30.
- Note that P.I. is much wider than C.I for mean.

Basics of Hypothesis Testing

Basic Steps for Testing Hypotheses

1. Determine the **null** hypothesis and the **alternative** hypothesis.
2. Collect **data** and summarize with a single number called a **test statistic**.
3. Determine how **unlikely** test statistic would be *if null hypothesis were true*.
4. Make a statistical **decision**.
5. Make a conclusion in **context**.

Step 1. Determine the hypotheses.

- **Null hypothesis**—hypothesis that says nothing is happening, status quo, no relationship, chance only, parameter equals a specific value (called "null value").
- **Alternative (research) hypothesis** — hypothesis is usually the reason data being collected; researcher suspects status quo belief is incorrect or that there is a relationship or change, or that the "null value" is not correct.

Class Input and Discussion

Share examples of hypothesis testing situations for your job:

- What was the question of interest?
- What were the null and alternative hypotheses?
- What kind of data did you have available to make a decision?

Example (analogy): A Jury Trial

If on a jury, must presume defendant is innocent unless enough evidence to conclude is guilty.

Null hypothesis: Defendant is *innocent*.
Alternative hypothesis: Defendant is *guilty*.

- Trial held because prosecution believes status quo of innocence incorrect.
- Prosecution collects evidence, like researchers collect data, in hope that jurors convinced such evidence extremely unlikely if assumption of innocence were true.

Step 2. Collect data and summarize with a test statistic.

Decision in hypothesis test based on single summary of data – the **test statistic.** Often this is a *standardized version of the point estimate*.

Step 3. Determine how unlikely test statistic would be if null hypothesis true.

If null hypothesis true, how likely to observe sample results of this magnitude or larger (in direction of the alternative) just by chance? … called p-value.

Step 4. Make a Statistical Decision.

Choice 1: *p*-value not small enough to convincingly rule out chance. We **cannot reject the null hypothesis** as an explanation for the results. There is **no statistically significant difference** or relationship evidenced by the data.

Choice 2: *p*-value small enough to convincingly rule out chance. We **reject the null hypothesis** and *accept* the alternative hypothesis. There is a **statistically significant difference** or relationship evidenced by the data.

How small is small enough?
Standard is **5%,** also called **level of significance**.

What Can Go Wrong:
The Two Types of Errors

Courtroom Analogy: *Potential choices and errors*

Choice 1: We believe we have enough evidence to conclude the defendant is guilty.
Potential error: Defendant is innocent; an innocent person falsely convicted and guilty party remains free.

Choice 2: We cannot rule out that defendant is innocent, so he or she is set free without penalty.
Potential error: Defendant is guilty; a criminal has been erroneously freed.

Choice 2 is usually seen as *more serious*.

Medical Analogy: *False Positive vs False Negative*

Tested for a disease; most tests not 100% accurate.
Null hypothesis: You do not have the disease.
Alternative hypothesis: You have the disease.

Choice 1: Medical practitioner thinks you have disease.
Test result strong enough to be "positive" for disease.
Potential error: You are healthy but told you're diseased.
Your test was a **false positive**.

Choice 2: Medical practitioner thinks you are healthy.
Test result weak enough to be "negative" for disease.
Potential error: You have disease but told you do not.
Your test was a **false negative**.

Which is *more serious*? Depends on disease and consequences.

The Two Types of Errors in Testing

DECISION MADE	TRUE STATE OF NATURE	
	Innocent, Healthy Null Hypothesis	Guilty, Diseased Alternative Hypothesis
Innocent Healthy Don't reject null hypothesis	Correct	Undeserved freedom False negative Type 2 error
Guilty Diseased Accept alternative hypothesis	Undeserved punishment False positive Type 1 error	Correct

- **Type 1 error** can only be made if the null hypothesis is actually true.
- **Type 2 error** can only be made if the alternative hypothesis is actually true.

Class Input and Discussion

In hypothesis tests for your job:

- What are the null and alternative hypotheses?
- What would be the conclusion in context if the null hypothesis were rejected?
- What would be a Type 1 error?
- What would be a Type 2 error?

Probabilities Associated with Errors

We can only specify the conditional probability of making a type 1 error, given that the null hypothesis is true. That probability is called the **level of significance**, usually 0.05.

Level of Significance and Type I Errors

- *If the null hypothesis is true, probability of making a type 1 error is equal to the level of significance, usually 0.05.*
- *If the null hypothesis is not true, a type 1 error cannot be made.*

Probabilities Associated with Errors

Type 2 Errors

- *A type 2 error is made if the alternative hypothesis is true, but you fail to choose it.*
- *A type 2 error can only be made if the alternative is true.*
- *The probability of doing that depends on which part of the alternative hypothesis is true, so computing the probability of making a type 2 error is not feasible unless a specific value in the alternative is specified.*

Probabilities Associated with Errors

The Power of a Test

- The **power** of a test is the probability of making the *correct decision* when the alternative hypothesis is true.
- Power = 1 – P(Type 2 error)
- If the population value falls close to the value specified in null hypothesis, then it can be difficult to get enough evidence from the sample to conclusively choose the alternative hypothesis, so *low power* if null and truth are close.
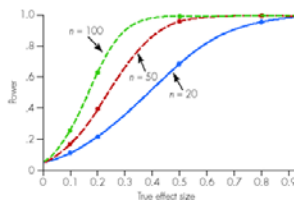

Type 2 Errors and Power

Three factors that affect probability of a type 2 error

1. **Sample size**; larger *n* reduces the probability of a type 2 error without affecting the probability of a type 1 error.
2. **Level of significance**; larger α reduces probability of a type 2 error by increasing the probability of a type 1 error.
3. **Actual value of the population parameter**; (not in researcher's control). Farther truth falls from null value (in direction of alternative), the lower the probability of a type 2 error.

Power curves, one-sided *t-test* for one mean; effect size = mean difference/s.d. (from *Mind On Statistics*)

| Sample Size	True Effect Size			
	0.1	0.2	0.5	0.8
n = 20	.11	.22	.695	.964
n = 50	.17	.40	.967	nearly 1
n = 100	.26	.63	.9996	nearly 1

Possible Errors and Level of Significance

In deciding whether to reject the null hypothesis consider the consequences of the two potential types of errors.

- If consequences of a type 1 error are very serious, then only reject null hypothesis if the *p*-value is very small.
- If type 2 error more serious, should be willing to reject null hypothesis with a moderately large *p*-value, 0.05 to 0.10.

Truth, decision, errors, probabilities
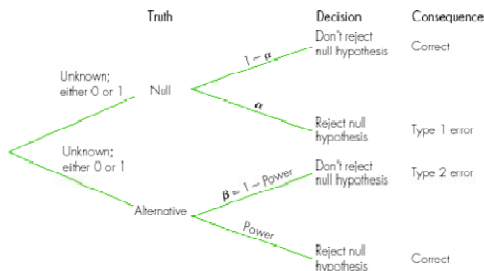
Real Importance versus Statistical Significance

A **statistically significant** relationship or difference *does not necessarily mean an important one*.

Whether results are statistically significant or not, it is helpful to examine a confidence interval so that you can determine the *magnitude* of the effect.

From **width** of the confidence interval, also learn how much **uncertainty** there was in sample results.

Role of Sample Size in Statistical Significance

If the sample size is **large enough**, almost **any null hypothesis can be rejected**.

There is almost always a *slight* relationship between two variables, or a difference between two groups, and if you *collect enough data, you will find it*.

No Difference versus No Statistically Significant Difference

• If the sample size is **too small**, an important relationship or difference can go **undetected**.

• In that case, we would say that the *power* of the test is *too low*.

• This is more likely to be a problem with your data than the other issue (statistical significance but not practical significance.

Finding Appropriate Sample Size and/or Power

• Specify what you think the truth is for the population.

• Specify the level of significance you plan to use (usually .05)

• Specify what power you want (probability of detecting the "truth" you specified above).

• Alternatively, specify what sample size you can afford, and compute power.

Resources and Example

• R Commander doesn't calculate power.

• Some good links at www.statpages.org

• Example (Carrie Austin): Compare mercury in Walker Creek delta and other areas. Suppose true means and std. deviations are:

	Walker Creek delta	Other locations
Mean	1.6	.5
Std Dev	1.1	.3

Link from statpages.org to *Power, sample size, for two groups* and click on first entry.

http://www.dssresearch.com/toolkit/sscalc/size_a2.asp

Result (next slide): Need *n* = 9 in each sample.

Used level of significance = .05 and power = .90.

	Walker Creek delta	Other locations
Mean	1.6	.5
Std Dev	1.1	.3

Hands-On Activity

Instructions on Activities Handout